

OPTIMASI HYPERPARAMETER PADA GRADIENT BOOSTED TREES MENGGUNAKAN BAYESIAN OPTIMIZATION

*Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika*



Oleh:

Arief Rachmatullah
09021181520022


JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2018

LEMBAR PENGESAHAN TUGAS AKHIR

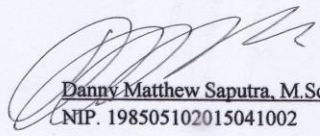
**OPTIMASI HYPERPARAMETER PADA GRADIENT BOOSTED
TREES MENGGUNAKAN BAYESIAN OPTIMIZATION**

Oleh :
ARIEF RACHMATULLAH
NIM : 09021181520022


Pembimbing I


Rifkie Primartha, M.T.
NIP. 197706012009121004

Inderalaya, Desember 2019
Pembimbing II,


Danny Matthew Saputra, M.Sc
NIP. 198505102015041002

Mengetahui,
Ketua Jurusan Teknik Informatika


Rifkie Primartha, M.T.
NIP. 197706012009121004

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Jumat tanggal 15 November 2019 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Arief Rachmatullah
N I M : 09021181520022
Judul : Optimasi *Hyperparameter* pada *Gradient Boosted Trees* Menggunakan *Bayesian Optimization*

1. Pembimbing I

Rifkie Primartha, M.T.
NIP. 19770601200912004

2. Pembimbing II

Danny Matthew Saputra, M.Sc
NIP. 198505102015041002

3. Penguji I

Novi Yusliani, M.T.
NIP. 198211082012122001

4. Penguji II

Desty Rodiah, M.T
NIP.

Mengetahui,
Ketua Jurusan Teknik Informatika

Rifkie Primartha, M.T.
NIP. 19770601200912004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Arief Rachmatullah
NIM : 09021181520022
Program Studi : Teknik Informatika
Judul Skripsi : Optimasi *Hyperparameter* pada *Gradient Boosted Trees*
Menggunakan *Bayesian Optimization*
Hasil Pengecekan Software *iThenticate/Turnitin* : 17%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 23 Desember 2019



Arief Rachmatullah
NIM. 09021181520022

Motto :

- *He who knows when he can fight and when he cannot will be victorious.*
–Sun Tzu
- *Perfectly balanced, as all things should be.*

Kupersembahkan karya tulis ini kepada :

- *Orang tuaku*
- *Keluarga besarku*
- *Sahabat dan teman seperjuanganku*
- *Fakultas Ilmu Komputer Universitas Sriwijaya*

HYPERPARAMETER OPTIMIZATION ON GRADIENT BOOSTED TREES
USING BAYESIAN OPTIMIZATION

By :
Arief Rachmatullah
09021181520022

ABSTRACT

Data mining is a technique to transform a collection of data into a knowledge. One of the factors that affect the accuracy of these data mining methods is hyperparameters which must be determined before training proses started. The data mining method that is very affected by hyperparameters is Gradient Boosted Trees. With a wrong hyperparameter configuration, the Gradient Boosted Tree model will be overfitting. One way that can be used to prevent this is to optimize Gradient Boosted Trees hyperparameters with hyperparameter optimization techniques and one of the popular *hyperparameter optimization* methods is *Bayesian Optimization*. Therefore, this study utilizes Bayesian Optimization to optimize Gradient Boosted Trees hyperparameter.

Keywords: Data Mining, *Hyperparameter Optimization*, Gradient Boosted Trees, Bayesian Optimization.

Pembimbing I

Rifkie Primatha, M.T
NIP. 197706012009121004

Inderalaya, Desember 2019
Pembimbing II,

Danny Matthew Saputra, M.Sc
NIP. 198505102015041002

Mengetahui,
Ketua Jurusan Teknik Informatika

Rifkie Primatha, M.T.
NIP. 197706012009121004

OPTIMASI HYPERPARAMETER PADA GRADIENT BOOSTED TREES
MENGUNAKAN BAYESIAN OPTIMIZATION

Oleh :
Arief Rachmatullah
09021181520022


ABSTRAK

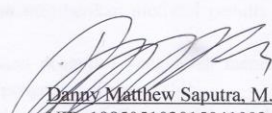
Data mining adalah teknik untuk mengubah kumpulan-kumpulan data yang banyak menjadi sebuah pengetahuan. Salah satu faktor yang mempengaruhi akurasi dari metode-metode *data-mining* yang ada ialah *hyperparameter* yang harus ditetapkan sebelum pelatihan data dimulai. Metode *data mining* yang sangat dipengaruhi dengan *hyperparameter* tersebut adalah *Gradient Boosted Trees*. Dengan pengaturan *hyperparameter* yang tidak tepat, maka model yang dihasilkan *Gradient Boosted Trees* akan menjadi *overfitting*. Salah satu cara yang dapat digunakan untuk mencegah hal tersebut adalah mengoptimasi *hyperparameter-hyperparameter Gradient Boosted Trees* dengan teknik *hyperparameter optimization* dan salah satu metode *hyperparameter optimization* yang populer adalah *Bayesian Optimization*. Maka dari itu, penelitian ini memanfaatkan *Bayesian Optimization* untuk mengoptimasi *hyperparameter Gradient Boosted Trees*.

Kata kunci: *Data Mining, Hyperparameter Optimization, Gradient Boosted Trees, Bayesian Optimization.*

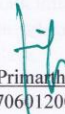
Pembimbing I

Inderalaya, Desember 2019
Pembimbing II,


Rifkie Primartha, M.T
NIP. 197706012009121004


Danny Matthew Saputra, M.Sc
NIP. 198505102015041002

Mengetahui,
Ketua Jurusan Teknik Informatika


Rifkie Primartha, M.T.
NIP. 197706012009121004

KATA PENGANTAR

Puji syukur kepada Allah SWT atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi syarat untuk menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang telah memberikan dukungan, bimbingan, dan motivasi kepada penulis untuk menyelesaikan tugas akhir ini, yaitu kepada:

1. Kedua orang tua penulis, Nursid Simenjaya, S.E. dan Zuhro Saidah, serta seluruh keluarga besarku yang selalu mendokan serta memberikan dukungan baik berupa moral maupun material.
2. Bapak Jaidan Jauhari, S.Pd, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Rifkie Primartha, S.T., M.T. selaku Ketua Jurusan Teknik Informatika dan sebagai pembimbing I Tugas Akhir yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
4. Bapak Danny Matthew Saputra, M.Sc selaku pembimbing II Tugas Akhir yang telah membimbing, mengarahkan dan memberikan motivasi kepada penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Bapak M. Fachrurrozi, S.Si., M.T selaku dosen pembimbing akademik, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan.
6. Ibu Novi Yusliani, M.T. selaku dosen penguji I yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
7. Ibu Desty Rodiah, M.T, M.T. selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
8. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
9. Kak Ricy dan Kak Hafez serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
10. Seluruh teman-teman jurusan Teknik Informatika dan teman-teman yang turut membantu dan mendukung penulis selama menempuh pendidikan di Fakultas Ilmu Komputer Universitas Sriwijaya.

11. Beserta semua pihak lain yang tidak dapat disebutkan satu persatu.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya.

Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, Desember 2019

Arief Rachmatullah

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO	v
<i>ABSTRACT</i>	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xv
BAB I	I-1
1.1 PENDAHULUAN	I-1
1.2 LATAR BELAKANG	I-1
1.3 RUMUSAN MASALAH	I-3
1.4 TUJUAN PENELITIAN	I-4
1.5 MANFAAT PENELITIAN	I-4
1.6 BATASAN MASALAH	I-4
1.7 SISTEMATIKA PENULISAN	I-5
1.8 KESIMPULAN	I-7
BAB II	II-1
2.1 PENDAHULUAN	II-1
2.2 <i>GRADIENT BOOSTED TREES</i>	II-1
2.3 <i>BAYESIAN OPTIMIZATION</i>	II-4
2.4 <i>CONFUSION MATRIX</i>	II-6
2.5 Rata-Rata dan Standar Deviasi	II-8
2.6 PENELITIAN LAIN YANG RELEVAN	II-9
2.7 KESIMPULAN	II-10
BAB III	III-1
3.1 PENDAHULUAN	III-1
3.2 PENGUMPULAN DATA	III-1
3.3 TAHAPAN PENELITIAN	III-2
3.3.1 Menentukan Ruang Lingkup dan Unit Penelitian	III-2

3.3.2 Menetapkan Dasar Teori yang Berkaitan dengan Penelitian.....	III-3
3.3.3 Menetapkan Kriteria Pengujian	III-3
3.3.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian....	III-6
3.3.5 Melakukan Pengujian Penelitian	III-6
3.3.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan Penelitian	III-9
3.4 METODE PENGEMBANGAN PERANGKAT LUNAK.....	III-10
3.4.1 Fase Insepsi.....	III-10
3.4.2 Fase Elaborasi.....	III-11
3.4.3 Fase Konstruksi	III-11
3.4.4 Fase Transisi	III-12
3.5 PENJADWALAN PENELITIAN.....	III-13
3.6 KESIMPULAN	III-19
BAB IV	IV-1
4.1 PENDAHULUAN.....	IV-1
4.2 <i>RATIONAL UNIFIED PROCESS</i>	IV-1
4.2.1. Fase Insepsi.....	IV-2
4.2.1.1. Pemodelan Bisnis	IV-2
4.2.1.2. Kebutuhan Sistem.....	IV-3
4.2.1.3. Analisis dan Desain	IV-5
4.2.1.3.1. Analisis Kebutuhan Perangkat Lunak.....	IV-5
4.2.1.3.2. Analisis Data	IV-6
4.2.1.3.3. Analisis Gradient Boosted Trees.....	IV-6
4.2.1.3.4. Analisis Optimasi Hyperparameter.....	IV-6
4.2.1.4. Desain Perangkat Lunak	IV-7
4.2.2. Fase Elaborasi.....	IV-18
4.2.2.1. Pemodelan Bisnis	IV-18
4.2.2.1.1. Perancangan Data	IV-18
4.2.2.2. Kebutuhan Sistem.....	IV-20
4.2.2.3. Diagram <i>Sequence</i>	IV-21
4.2.3. Fase Konstruksi	IV-26
4.2.3.1. Kebutuhan Sistem.....	IV-26
4.2.3.2. Diagram Kelas	IV-26
4.2.3.3. Implementasi	IV-27
4.2.4. Fase Transisi	IV-32
4.2.4.1. Pemodelan Bisnis	IV-32
4.2.4.2. Kebutuhan Sistem.....	IV-32
4.2.4.3. Rencana Pengujian	IV-33
4.2.4.4. Implementasi	IV-35
4.3 KESIMPULAN	IV-42
BAB V.....	V-1

5.1 PENDAHULUAN.....	V-1
5.2 DATA HASIL PERCOBAAN/PENELITIAN.....	V-1
5.2.1 Percobaan.....	V-1
5.2.2 Data Hasil Pengujian Akurasi Gradient Boosted Trees.....	V-1
5.2.2 Data Hasil Pengujian Akurasi Gradient Boosted Trees + Bayesian Optimization	V-3
5.3 ANALISIS HASIL PENELITIAN.....	V-5
5.4 KESIMPULAN	V-7
BAB VI	VI-1
6.1 PENDAHULUAN.....	VI-1
6.2 KESIMPULAN	VI-1
6.3 SARAN.....	VI-2

DAFTAR TABEL

Tabel II-1. <i>Confusion Matrix</i>	II-7
Tabel III – 1. Rancangan perbandingan rata-rata akurasi dan standar deviasi ..	III-7
Tabel III – 2. Rancangan Tabel Hasil Pengujian Akurasi Untuk <i>Dataset Banknote Authentication</i>	III-8
Tabel III – 3. Rancangan Tabel Hasil Pengujian Akurasi Untuk <i>Dataset Staglog Heart</i>	III-9
Tabel III – 4. Rancangan Tabel Perbandingan Rata-Rata Akurasi dan Standar Deviasi	III-9
Tabel III – 5. Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure (WBS)</i>	III-14
Tabel IV – 1 Tabel Kebutuhan Fungsional	IV-4
Tabel IV – 2 Tabel Kebutuhan Non-Fungsional	IV-5
Tabel IV – 3. Tabel Definisi Aktor	IV-8
Tabel IV – 4. Tabel Definisi <i>Use Case</i>	IV-9
Tabel IV – 5. Tabel Skenario <i>Use Case Training Data</i>	IV-10
Tabel IV – 6. Tabel <i>Use Case Testing Data</i>	IV-11
Tabel IV – 7. Tabel <i>Use Case</i> Melakukan Optimasi <i>Hyperparameter</i>	IV-12
Tabel IV – 8. Tabel <i>Use Case</i> Menyimpan <i>Model</i>	IV-13
Tabel IV – 9. Tabel <i>Use Case</i> Melakukan <i>Data Mining</i> Klasifikasi	IV-14
Tabel IV – 10. Tabel Implementasi Kelas	IV-28
Tabel IV – 11. Tabel Rencana Pengujian <i>Use Case Training Data</i>	IV-33
Tabel IV – 12. Tabel Rencana Pengujian <i>Use Case Testing Data</i>	IV-34
Tabel IV – 13. Tabel Rencana Pengujian <i>Use Case</i> Melakukan <i>Hyperparameter Optimization</i>	IV-34
Tabel IV – 14. Tabel Rencana Pengujian <i>Use Case</i> Menyimpan <i>Model</i>	IV-35
Tabel IV – 15. Tabel Rencana Pengujian <i>Use Case</i> Melakukan <i>Data Mining</i> Klasifikasi	IV-35
Tabel IV – 16. Tabel Pengujian <i>Use Case Training Data</i>	IV-36

Tabel IV – 17. Tabel Pengujian <i>Use Case Testing Data</i>	IV-37
Tabel IV – 18. Tabel Pengujian <i>Use Case Melakukan Hyperparameter Optimization</i>	IV-38
Tabel IV – 19. Tabel Pengujian <i>Use Case Menyimpan Model</i>	IV-39
Tabel IV – 20. Tabel Pengujian <i>Use Case Melakukan Data Mining Klasifikasi</i>	IV-40
Tabel V – 1. Tabel Hasil Pengujian Akurasi <i>Gradient Boosted Trees</i> pada <i>Dataset Climate Model Simulation Crashes</i>	V-1
Tabel V – 2. Tabel Hasil Pengujian Akurasi <i>Gradient Boosted Trees</i> pada <i>Dataset Banknote Authentication</i>	V-2
Tabel V – 3. Tabel Hasil Pengujian Akurasi <i>Gradient Boosted Trees</i> pada <i>Dataset Staglog Heart</i>	V-3
Tabel V – 4. Tabel Hasil Pengujian Akurasi <i>Gradient Boosted Trees + Bayesian Optimization</i> pada <i>Dataset Climate Model Simulation Crashes</i>	V-3
Tabel V – 5. Tabel Hasil Pengujian Akurasi <i>Gradient Boosted Trees + Bayesian Optimization</i> pada <i>Dataset Banknote Authentication</i>	V-4
Tabel V – 6. Tabel Hasil Pengujian Akurasi <i>Gradient Boosted Trees + Bayesian Optimization</i> pada <i>Dataset Staglog Heart</i>	V-5
Tabel V – 7. Tabel Perbandingan Rata-Rata Akurasi dan Standar Deviasi.....	V-6

DAFTAR GAMBAR

Gambar II – 1. Diagram alur <i>boosting</i>	II-2
Gambar III – 1. Alur kerja <i>Gradient Boosted Trees + Bayesian Optimization</i> .	III-2
Gambar III – 2. Alur pengujian penelitian	III-5
Gambar IV – 1. Diagram <i>Use Case</i>	IV-8
Gambar IV – 2. Diagram Aktivitas <i>Training Data</i>	IV-16
Gambar IV – 3. Diagram Aktivitas <i>Testing Data</i>	IV-16
Gambar IV – 4. Diagram Aktivitas Melakukan Optimasi <i>Hyperparameter</i> ...	IV-17
Gambar IV – 5. Diagram Aktivitas Mengklasifikasi menggunakan <i>Gradient Boosted Trees</i>	IV-17
Gambar IV – 6. <i>Wireframe MainFrame</i>	IV-19
Gambar IV – 7. <i>Wireframe</i> Menu Klasifikasi	IV-20
Gambar IV – 8. <i>Sequence Diagram Training Data</i>	IV-22
Gambar IV – 9. <i>Sequence Diagram Testing Data</i>	IV-23
Gambar IV – 10. <i>Sequence Diagram</i> Melakukan Optimasi <i>Hyperparameter</i>	IV-24
Gambar IV – 11. <i>Sequence Diagram</i> Menyimpan <i>Model</i>	IV-25
Gambar IV – 12. <i>Sequence Diagram</i> Mengklasifikasi menggunakan <i>Gradient Boosted Trees</i>	IV-25
Gambar IV – 13. Diagram Kelas.....	IV-27
Gambar IV – 14. Antarmuka <i>MainFrame</i> Perangkat Lunak	IV-30
Gambar IV – 15. Antarmuka Menu Klasifikasi Perangkat Lunak.....	IV-31

DAFTAR LAMPIRAN

LAMPIRAN 1. Kode Program	L-1
--------------------------------	-----

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini dibahas mengenai masalah yang melatarbelakangi penelitian ini, rumusan masalah, tujuan dari penelitian ini, manfaat dari penelitian ini serta batasan-batasan masalah yang terdapat pada penelitian ini. Pada bab ini juga akan dipaparkan penjelasan umum mengenai keseluruhan penelitian.

1.2 Latar Belakang

Data mining adalah cara untuk mengubah kumpulan-kumpulan data yang banyak menjadi sebuah pengetahuan (Han, Kamber and Pei, 2011). Saat ini *data mining* menjadi topik yang banyak dibicarakan karena data yang dapat dihasilkan dari satu orang bisa sangat banyak. Untuk mengubah kumpulan-kumpulan data yang dihasilkan seseorang tersebut menjadi sebuah pengetahuan, maka diperlukan beberapa teknik dan metode-metode tertentu sehingga dapat diambil suatu keputusan.

Metode-metode *data mining* modern saat ini melibatkan parameter-parameter yang harus ditetapkan sebelum proses pelatihan data dilakukan, berbeda dengan parameter lain yang dapat berubah sesuai dengan proses pelatihan data. Parameter tersebut biasa disebut dengan *hyperparameter* (Rijn and Hutter, 2018). *Hyperparameter* memiliki peran yang penting dikarenakan pengaturan *hyperparameter* dapat mempengaruhi *performance* dari suatu metode *data mining* (Rijn and Hutter, 2018). Beberapa metode data mining yang melibatkan beberapa

hyperparameter tersebut diantaranya adalah *Support Vector Machine (SVM)*, *Random Forest*, dan *Gradient Boosted Trees*.

Gradient Boosted Trees merupakan algoritma pengklasifikasi yang menggunakan teknik boosting, yaitu terdiri dari kumpulan beberapa model dasar yang akan dikombinasikan sehingga menjadi model akhir yang akurat (Krauss, Do and Huck, 2016). Dengan teknik tersebut, dapat dibuktikan bahwa algoritma *Gradient Boosted Trees* dapat menghasilkan model yang lebih akurat dibandingkan dengan algoritma pengklasifikasi yang lainnya (Zhang and Haghani, 2015). Pada *Gradient Boosted Trees*, yang dijadikan sebagai model dasar adalah *decision tree* dan pada *Gradient Boosted Trees* juga terdapat beberapa *hyperparameter* yang akan dikonfigurasi terlebih dahulu. Konfigurasi *hyperparameter* pada *Gradient Boosted Trees* menjadi salah satu hal yang penting karena salah satu kelemahan pada *Gradient Boosted Trees* yaitu mudahnya terjadi *overfitting* yang disebabkan karena konfigurasi *hyperparameter* yang tidak tepat. Maka dari itu perlu dilakukan optimasi *hyperparameter* pada *Gradient Boosted Trees*.

Telah dilakukannya penelitian terkait optimasi *hyperparameter*, salah satunya yaitu penggunaan algoritma genetika untuk mengoptimasi *hyperparameter* pada algoritma *Random Forest* (Elyan and Gaber, 2017). Hasil penelitian mereka menunjukkan bahwa dengan mengoptimasi *hyperparameter* menggunakan algoritma genetika, model yang dihasilkan dari *Random Forest* menjadi lebih akurat.

Selain algoritma genetika, *Bayesian optimization* juga dapat digunakan untuk mengoptimasi *hyperparameter*. *Bayesian Optimization* adalah metode optimasi *global* untuk fungsi *black-box* (Snoek, Rippel and Adams, 2015). *Bayesian optimization* menggunakan *acquisition function* untuk menyeimbangkan antara eksplorasi dan eksploitasi pada penentuan nilai *hyperparameter* berikutnya. Menurut Jones, algoritma ini memiliki performa yang lebih baik dari pendekatan optimasi yang lainnya (J. Snoek, Larochelle and Adams, 2012).

Berdasarkan hal diatas, maka dalam penelitian ini akan dilakukan optimasi *hyperparameter* pada *Gradient Boosted Trees* menggunakan *Bayesian Optimization* dengan harapan dapat mengurangi resiko *overfitting* sehingga meningkatkan akurasi dari *Gradient Boosted Trees*.

1.3 Rumusan Masalah

Berdasarkan latar belakang diatas maka rumusan masalah dalam penelitian ini adalah bagaimana perbandingan akurasi *Gradient Boosted Trees* jika dilakukan optimasi *hyperparameter* menggunakan *Bayesian Optimization*.

Untuk menjawab rumusan masalah di atas maka rumusan masalah tersebut akan diuraikan menjadi tiga pertanyaan penelitian (*Research Question*), yaitu:

1. Bagaimana cara kerja *Gradient Boosted Trees*?
2. Bagaimana cara kerja *Bayesian Optimization*?
3. Bagaimana cara membandingkan akurasi antara *Gradient Boosted Trees* yang telah dilakukan optimasi *hyperparameter* menggunakan

Bayesian Optimization dan *Gradient Boosted Trees* tanpa optimasi *hyperparameter*?

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Membentuk model *Gradient Boosted Trees*.
2. Melakukan optimisasi *hyperparameter* pada algoritma *Gradient Boosted Trees* menggunakan *Bayesian Optimization*.
3. Membandingkan akurasi antara *Gradient Boosted Trees* yang telah dilakukan optimisasi *hyperparameter* menggunakan *Bayesian Optimization* dan *Gradient Boosted Trees* tanpa optimisasi *hyperparameter*.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah:

1. Membantu memudahkan peneliti-peneliti lain dalam melakukan optimisasi *hyperparameter*.

1.6 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. *Dataset* yang akan digunakan diambil dari *UCI Machine Learning Repository*, yaitu *Banknote Authentication*, *Staglog Heart*, dan *Climate Model Simulation Crashes*.
2. *Hyperparameter* yang akan dioptimasi adalah banyak iterasi dan *learning rate*.
3. Pengukuran akurasi dihitung menggunakan confusion matriks

1.7 Sistematika Penulisan

Penelitian ini disusun sedemikian rupa berdasarkan sistematika penulisan di bawah ini:

BAB I. PENDAHULUAN

Pada bab pendahuluan dipaparkan mengenai latar belakang penelitian, rumusan masalah pada penelitian ini, tujuan penelitian, manfaat penelitian, batasan masalah di dalam penelitian atau ruang lingkup penelitian serta sistematika penulisan.

BAB II. TINJAUAN PUSTAKA

Selanjutnya pada tinjauan pustaka dilakukan pembahasan mengenai teori-teori yang dijadikan sebagai dasar atau landasan pada penelitian ini, seperti algoritma *Gradient Boosted Trees*, algoritma *Bayesian Optimization*, cara mengukur tingkat akurasi pada suatu model klasifikasi serta beberapa kajian literatur mengenai penelitian lain yang relevan pada penelitian ini.

BAB III. METODOLOGI PENELITIAN

Setelah dilakukan tinjauan pustaka, dilanjutkan dengan metodologi penelitian yang berisi tentang bahasan dan tahapan yang dilaksanakan pada penelitian ini. Setiap rencana-rencana tahapan penelitian akan dipaparkan secara rinci dan mengacu pada suatu kerangka kerja. Dan pada akhir metodologi penelitian akan dipaparkan rancangan manajemen proyek pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Kemudian pada bab IV yaitu pengembangan perangkat lunak, akan dilakukan pembahasan analisis serta perancangan pengembangan perangkat lunak yang akan digunakan sebagai alat penelitian dalam penelitian ini. Bab ini dimulai dengan pengumpulan *user requirements* dan analisis kebutuhan, rancangan serta konstruksi perangkat lunak yang akan dikembangkan serta pengujian untuk memastikan semua kebutuhan pengembangan perangkat lunak sesuai dengan kebutuhan. Penyusunan pada bab ini memiliki kerangka penulisan dengan fase-fase dan elemen-elemen pengembangan perangkat lunak bersifat berorientasi objek.

BAB V. HASIL DAN ANALISA PENELITIAN

Lalu pada bab hasil dan analisa penelitian dijelaskan hasil pengujian berdasarkan langkah-langkah yang telah direncanakan pada metodologi penelitian. Tabel-tabel hasil pengujian serta analisisnya diapaparkan sebagai landasan untuk menarik kesimpulan yang akan diambil dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Kemudian pada kesimpulan dan saran akan dipaparkan kesimpulan-kesimpulan yang dapat diambil dari uraian-uraian pada penelitian ini beserta dengan saran-saran yang diharapkan berguna dalam optimasi pada *Gradient Boosted Trees* menggunakan *Bayesian Optimization*.

1.8 Kesimpulan

Pada bab pendahuluan ini, telah diuraikan secara umum mengenai isu dan masalah-masalah yang melatarbelakangi penelitian ini. Dan pada bab ini juga telah dipaparkan rumusan masalah pada penelitian ini, tujuan dari penelitian ini, manfaat dari penelitian ini, batasan-batasan masalah pada penelitian ini serta sistematika penulisan pada penelitian ini sehingga penelitian ini mempunyai landasan-landasan yang cukup kuat.

DAFTAR PUSTAKA

- Blagus, R. and Lusa, L. (2016) 'Gradient boosting for high-dimensional prediction of rare events', *Computational Statistics and Data Analysis*. Elsevier B.V. doi: 10.1016/j.csda.2016.07.016.
- Dewancker, I. *et al.* (2015) 'Bayesian Optimization Primer', pp. 2–5.
- Elyan, E. and Gaber, M. M. (2017) 'A genetic algorithm approach to optimising random forests applied to class engineered data', *Information Sciences*. Elsevier Inc., 384, pp. 220–234. doi: 10.1016/j.ins.2016.08.007.
- Han, J., Kamber, M. and Pei, J. (2011) *Data Mining Concepts and Techniques*. 3rd edn. Elsevier.
- Krauss, C., Do, X. A. and Huck, N. (2016) 'Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500', *European Journal of Operational Research*. Elsevier B.V., 259, pp. 689–702. doi: 10.1016/j.ejor.2016.10.031.
- Lizotte, D. J. (2008) *Practical Bayesian Optimization*. University of Alberta.
- Lorenz, R. *et al.* (2016) 'Stopping criteria for boosting automatic experimental design using real-time fMRI with Bayesian optimization *', pp. 1–8.
- Morar, M. T., Knowles, J. and Sampaio, S. (2017) 'Initialization of Bayesian Optimization Viewed as Part of a Larger Algorithm Portfolio', in *Data Science meets Optimization Workshop: CEC2017 & CPAIOR 2017: DSO 2017.*, p. 6.
- Natekin, A. and Knoll, A. (2013) 'Gradient boosting machines , a tutorial', 7(December), p. 21. doi: 10.3389/fnbot.2013.00021.
- Pressman, R. S. (2005) *Software Engineering*.
- Rijn, J. N. Van and Hutter, F. (2018) 'Hyperparameter Importance Across Datasets', in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2367–2376.
- Shultz, T. R. and Fahlman, S. E. (2017) *Encyclopedia of Machine Learning and Data Mining*. doi: 10.1007/978-1-4899-7687-1.
- Snoek, B. J., Larochelle, H. and Adams, R. P. (2012) 'PRACTICAL BAYESIAN OPTIMIZATION OF MACHINE LEARNING', in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2*, pp. 1–12.
- Snoek, J., Larochelle, H. and Adams, R. P. (2012) 'Practical Bayesian Optimization of Machine Learning Algorithms', in *NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2*. Lake Tahoe, Nevada, pp. 2951–2959.

- Snoek, J., Rippel, O. and Adams, R. P. (2015) 'Scalable Bayesian Optimization Using Deep Neural Networks', in *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, pp. 2171–2180.
- Zhang, Y. and Haghani, A. (2015) 'A gradient boosting method to improve travel time prediction', *Transportation Research Part C: Emerging Technologies*. Elsevier Ltd, 58, pp. 308–324. doi: 10.1016/j.trc.2015.02.019.